Data Analysis 4

Summary Stats

General Maths - Preliminary

Name: _____

PRELIMINARY CAPACITY MATRIX – GENERAL MATHEMATICS

TOPIC: Data Analysis 4 – Summary Statistics 2 weeks

CONTENT	CAPACITY BREAKDOWN!	DONE IT!!!!	GOT IT!!!!!	ON MY WAY!	WORKING ON IT!	HELP!!!!
1. The mean a. Calculation with grouped data b. Calculating scores given the mean c. Investigating the effect on the mean when adding to scores etc 2. Determination of the median and mode(s) of a data set	Skillsheets 10.1-10.4 The Mean, Mode, Median, Range W/S Changes to the Mean W/S Netball Team W/S Ex 10A DIY Mean excel sheet Ex 10C					
3. Informal description of standard deviation as a measure of the spread of data in relation to the mean 4. Determination of the population standard deviation and sample standard deviation	Ex 10B (do a and c of every question) HSC comparison W/S					
5. Selection and use of the appropriate statistic to describe features of a data set	Ex 10D					

Your say!

What was the most important thing you learned?_____

What was something new you learnt? _____

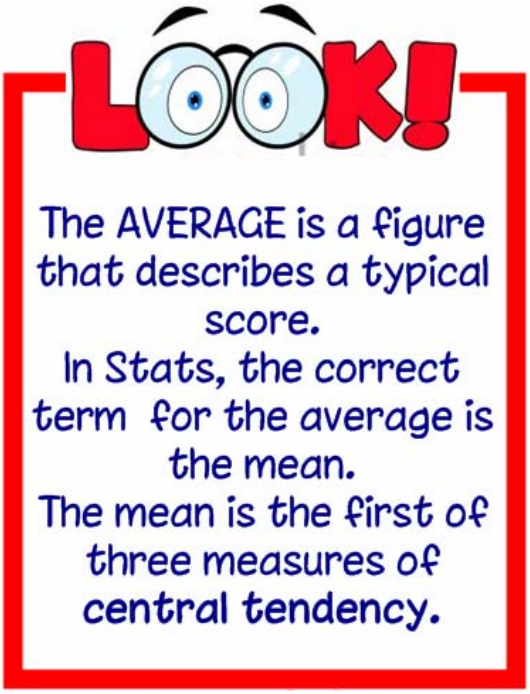
What part(s) of this topic will you need to work on? _____

Review...

Mean

The statistical symbol for mean is $\bar{x} = \frac{\sum x}{n}$ (the sum of the scores divided by the number of terms.

eg Calculate the mean of the scores 17, 16, 13, 15, 16, 20, 12, 15



Calculating the mean of a frequency distribution.

1. Add an extra column to the frequency distribution, which is the $f.x$ (frequency x score) column.
2. Total the column.
3. Divide the total by the sum of the frequency column.

$$x = \frac{\Sigma f x}{\Sigma f}$$

eg Calculate the frequency table then calculate the mean.

Score (x)	Frequency (f)	$f \times x$
4	3	
5	7	
6	11	
7	13	
8	10	
9	6	
	$\Sigma f =$	$\Sigma f \times x =$

Calculating the mean of grouped data.

1. Add an extra column to the frequency distribution, which is the $f.x$ (frequency x score) column.
2. Calculate the $f.x$ column by multiplying the class centre by the frequency.
3. Total the column.
4. Divide the total by the sum of the frequency column.

$$\bar{x} = \frac{\Sigma f x}{\Sigma f}$$

eg Complete the frequency distribution table and use it to estimate the mean of the distribution.

Class	Class centre (x)	Frequency (f)	$f \times x$
25–29		4	
30–34		9	
35–39		13	
40–44		12	
45–50		7	
		$\Sigma f =$	$\Sigma f \times x =$

Median

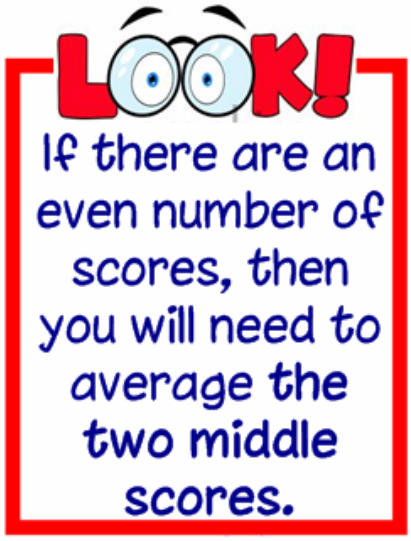
Median is the middle score when the scores are in consecutive order.

Median vs Mean

When is it advantageous to use the median over the mean?

eg Analyse the typical house prices in an area. On one particular day, five houses are sold for the following prices:

$374 000, $348 900, $361 000, $411 000, $1 243 000



eg Calculate the median of the scores:

a) 3, 5, 7, 4, 4, 7, 9, 1, 6

b) 13, 13, 16, 12, 19, 17, 21, 18

Calculating the median from a frequency distribution.

1. Add a cumulative frequency column to the frequency distribution.
2. Halve the total number of scores (h).
3. Find the two values in the c.f column that are either side of h.
4. If there is an even number of scores, you will need to find the average of h and h+1.

eg Calculate the median for the frequency distribution.

Score	Frequency
34	3
35	8
36	12
37	9
38	8
39	5

IF THERE IS GROUPED DATA, THE MEDIAN IS ESTIMATED FROM THE OGIVE.

Mode

There are times when both the mean and median are inappropriate measures of the typical score in the data.

Investigation:

eg A clothing store is assessing the sizes that sell the most. The following data Is collected:

8 12 14 12 16 10 12 14 16 18

8 14 12 14 12 12 18 16 12 14

Calculate the mean and median.

THE MODE IS THE SCORE WITH THE LARGEST FREQUENCY

eg Find the mode of the following set of scores:

4, 5, 9, 4, 6, 8, 4, 8, 7, 6, 5, 4

STANDARD DEVIATION

The most commonly used measure of spread is the **STANDARD DEVIATION.**

The standard deviation is a measure of how much a typical score in a data set differs from the mean. The larger the deviation, the greater the spread of scores from the mean. The smaller the deviation, the more consistent the scores are.

There are two standard deviation functions on your calculator. The first, σ_n is the **population standard deviation**. This function is used when the statistical analysis is conducted on the entire population.

When the statistical analysis is done using a sample of the population, then the **sample deviation function** is used. The symbol for this is σ_{n-1}.

eg A survey of cricket scores is collated. The results are

65 82 47 78 108 94 60 79 88 91

50 73 68 95 83 76 79 72 69 97

Calculate the mean and standard deviation.

Comment on the reliability of the batsman.